STAGE-DEPENDENT FUZZY LOSS FUNCTION IN MULTISTAGE CLASSIFIER – RESULTS OF SIMULATIONS

Robert Burduk

Wroclaw University of Technology, Chair of Systems and Computer Networks,
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

robert.burduk@pwr.wroc.pl (Robert Burduk)

Abstract

The work deals with a recognition problem using a probabilistic-fuzzy model. The model is based on the notion of fuzzy random variable and also on the Bayesian theory. The Bayesian hierarchical classifier is based on a decision-tree scheme. For given tree skeleton and features to be used, the optimal (Bayes) decision rules (strategy) at each non-terminal node are presented. The globally optimal Bayes strategy (which minimizes the overall error probability) has been calculated for stage-dependent loss function. This fuzzy loss function means that the loss depends on the stage at which misclassification is made. The loss function in our case is fuzzy-valued and is described by a fuzzy triangular or trapezoidal number. In order to rank fuzzy mean values, we have selected the subjective method defined by Campos and Gonzalez. This method is based on the λ-average valued of a fuzzy number where λ parameter is a subjective degree of optimism-pessimism. In the end, some results of simulation investigations of this case of pattern recognition are presented. This results presented influence of parameter λ on separation point for decision regions at the first stage. This paper contribute to a better understanding of the impact of the choice of a fuzzy numbers which describe stage-dependent loss function in multistage classifier.

Keywords: Fuzzy loss function, Bayes rule, Multistage classifier.

Presenting Author's biography

Robert Burduk received an MSc degree in Electronic Engineering and a PhD degree in Computer Engineering from the Wroclaw University of Technology in 1998 and 2003, respectively. His PhD dissertation dealt with multistage pattern recognition with fuzzy-probabilistic model. Currently he is a lecturer at the Wroclaw University of Technology. He teaches pattern recognition, data mining and databases.

ISBN 978-3-901608-32-2

1 Introduction

This paper deals with a recognition problem, which – assuming a probabilistic model with a full information – values of a loss function are assumed to be fuzzy numbers. The class of the fuzzy-valued loss functions is definitely much wider than the class of the real-valued ones. This fact reflects the richness of the fuzzy expected loss approach to describe the consequences of incorrect classification as opposed to the real-valued approach. For this reason, several studies have previously described decision problems in which values assessed to the consequences of decision are assumed to be fuzzy [1,5,6,11]. These papers describe only single-stage decision problems.

This paper deals with a multistage recognition problem We will consider the so-called Bayesian hierarchical classifier [8,9]. In this recognition problem the decision as to the membership of an object into a class is not a single activity but is the result of a more or less complex decision process. The mechanics of classification can be described by means of a tree, in which the terminal nodes represent pattern classes, i.e. final classification, and the interior nodes denote groups of classes. In particular, the root-node represents the entire set of classes into which a pattern may be classified. This model has been formulated so that, on the one hand, the existence of fuzzy loss function representing the preference pattern of the decision maker can be established; while, on the other hand, a priori probabilities of classes and class-conditional probability density functions can be given.

In the further part, after the introduction of necessary symbols, we will calculate the set of optimal recognition algorithms for internal nodes, minimizing the global quality indicator. As a criterion of optimality we will assume the mean value of the fuzzy loss function (risk), which values depends on the stage of the decision tree, on which an error has occurred. The presented algorithm will be illustrated by a simulation investigation in which crisp method for ranking fuzzy numbers was applied.

2 The multistage recognition task

The synthesis of multistage classifier is a complex problem. It involves specification of the following components:

- the decision logic, i.e. hierarchical ordering of classes,
- feature used at each stage of decision,
- the decision rules (strategy) for performing the classification.

This paper is devoted only to the last problem. This means that we will deal only with the presentation of decision algorithms, assuming that both the tree skeleton and the feature used at each non-terminal node have been specified.

Let us consider a pattern recognition problem, in which the number of classes is equal to M. Let us assume that classes were organized in (N+1) horizontal decision tree. Let us number all nodes of the constructed decision-tree with consecutive numbers of 0, 1, 2,..., reserving 0 for the root-node and let us assign numbers of classes from the $\mathcal{M}=\{1,2,...,M\}$ set to terminal nodes so that each one of them is labeled with the number of the class which is connected with that node. This allows the introduction of the following notation [8,9]:

- $\mathcal{M}(n)$ – the set of numbers of nodes, which distance from the root is n, $n=0,1,2,...,N$. In particular $\mathcal{M}(0)=\{0\}$, $\mathcal{M}(N)=\mathcal{M}$,
- $\overline{\mathcal{M}}=\bigcup_{n=0}^{N-1}\mathcal{M}(n)$ – the set of interior node numbers (non terminal),
- $\mathcal{M}_i \subseteq \mathcal{M}(N)$ – the set of class numbers attainable from the i-th node ($i\in\overline{\mathcal{M}}$),
- $\mathcal{M}^i$ – the set of numbers of immediate descendant nodes i ($i\in\overline{\mathcal{M}}$),
- m_i – number of immediate predecessor of the i–th node (i≠0).

We will continue to adopt the probabilistic model of the recognition problem, i.e. we will assume that the class number of the pattern being recognized $j_N\in\mathcal{M}(N)$ and its observed features x are realizations of a couple of random variables $\boldsymbol{J}_N$ and **X**. Complete probabilistic information denotes the knowledge of a priori probabilities of classes:

$$p(j_N)=P(J_{\boldsymbol{N}}=j_N),\quad j_N\in\mathcal{M}(N) \tag{1}$$

and class-conditional probability density functions:

$$f_{j_N}(x)=f(x/j_N),\quad x\in X,\quad j_N\in\mathcal{M}(N). \tag{2}$$

Let

$$x_i\in\mathcal{X}_i\subseteq\mathcal{R}^{d_i},\ d_i\le d,\ i\in\mathcal{M} \tag{3}$$

denote vector of features used at the i-th node, which have been selected from the vector x.

Our target now is to calculate the so-called multistage recognition strategy $\pi_N=\{\psi_i\}_{i\in\overline{\mathcal{M}}}$., that is the set of recognition algorithms in the form:

$$\Psi_i:\mathcal{X}_i\rightarrow\mathcal{M}^i,\ i\in\overline{\mathcal{M}}\ . \tag{4}$$

Formula (4) is a decision rule (recognition algorithm) used at the i-th node which maps observation subspace to the set of immediate descendant nodes of the *i*-th node.

Let $L(i_N, j_N)$ denote the loss incurred when objects of the class j_N is assigned to the class i_N ($i_N, j_N \in \mathcal{M}(N)$). Our aim is to minimize the mean risk, that is the mean value of the fuzzy loss function [6]:

$$\tilde{R}^*(\pi_N{}^*) = \min \underset{\boldsymbol{I}_N, \boldsymbol{J}_N}{E} [L(\boldsymbol{I}_N, \boldsymbol{J}_N)] = \tag{5}$$

$$= \min \underset{\boldsymbol{X}, \boldsymbol{J}_N}{E} [L(\Psi(\boldsymbol{X}), \boldsymbol{J}_N)]$$

We will refer to the π_N^* strategy as the globally optimal *N*-stage recognition strategy.

The fuzzy-valued loss function is described by triangular or trapezoidal fuzzy number. The trapezoidal fuzzy number is characterized by means of a membership function:

$$\mu_A(x)_{Tr} = \begin{cases} 0 & x \le a \\ f_A(x) & a \le x \le b \\ 1 & b \le x \le c \\ g_A(x) & c \le x \le d \\ 0 & x \ge d \end{cases} \tag{6}$$

and triangular fuzzy number is characterized by means of a membership function:

$$\mu_A(x)_T = \begin{cases} 0 & a \le x \\ f_A(x) & a \le x \le b \\ g_A(x) & b \le x \le c \\ 0 & x \ge c \end{cases} \tag{7}$$

where f_A and g_A are continuous functions, f_A is increasing (from 0 to 1), g_A is decreasing (from 1 to 0).

In order to rank fuzzy mean values, we have selected the subjective method defined by Campos and Gonzalez [4]. This method is based on the λ-average valued of a fuzzy number, which is defined for $\tilde{A}$ as the real number given by

$$V_S^\lambda(\tilde{A}) = \int_0^1 [\lambda a_{\alpha 2} + (1-\lambda) a_{\alpha 1}] dS(\alpha) \tag{8}$$

where $\tilde{A}_\alpha = [a_{\alpha 1}, a_{\alpha 2}]$, $\lambda \in [0,1]$ and S being an additive measure on $Y \subset [0,1]$.

The λ parameter is a subjective degree of optimism-pessimism. In a loss context, λ=0 reflects the highest optimism, while λ=1 reflects the highest pessimism. Then, the λ-ranking method used for comparing fuzzy numbers is given by:

$$\tilde{A} \ge \tilde{B} \Leftrightarrow V_S^\lambda(\tilde{A}) \ge V_S^\lambda(\tilde{B}) \,. \tag{9}$$

This λ-average ranking method extends some well-known ranking functions [3,12]. One of the most relevant characteristics of the ranking method based on the function V_S^λ is its feasibility, which is due to the following reason: when we apply V_S^λ on the fuzzy expected value of an integrably bounded fuzzy random variable the computation of this fuzzy expected value is not required. The λ-average value of a fuzzy expected value is reduced to the expected value of a classical random variable, namely, the composition of V_S^λ and the fuzzy variable [10].

3 The recognition algorithm for stage-dependent fuzzy loss function

Let us assume now

$$\tilde{L}(i_N, j_N) = \tilde{L}^s_{\mathrm{d}(w)} \tag{10}$$

where w is the first common predecessor of the nodes i_N and j_N ($i_N, j_N \in \mathcal{M}(N)$). So defined fuzzy loss function means that the loss depends on the stage at which misclassification is made. Stage-dependent fuzzy loss function for the three-stage binary classifier are presented in Fig. 1.

Putting (5) into (4) we obtain the optimal (Bayes) strategy, which decision rules are as follows:

$$\Psi^*_{i_n}(x_{i_n}) = i_{n+1} \tag{11}$$

$$(\tilde{L}^s_{\mathrm{d}(i_n)} - \tilde{L}^s_{\mathrm{d}(i_{n+1})}) p(i_{n+1}) f_{i_{n+1}}(x_{i_n}) +$$

$$+ \sum_{j_{n+2} \in \mathcal{M}^{i_{n+1}}} [(\tilde{L}^s_{\mathrm{d}(i_{n+1})} - \tilde{L}^s_{\mathrm{d}(j_{n+2})}) q^*(j_{n+2} / i_{n+1}, j_{n+2}) \times$$

$$\times p(j_{n+2}) f_{j_{n+2}}(x_{i_n}) +$$

$$+ \ldots + \tilde{L}^s_{\mathrm{d}(j_{N-1})} \sum_{j_N \in \mathcal{M}^{j_{N-1}}} [q^*(j_N / i_{n+1}, j_N) \times$$

$$\times p(j_N) f_{j_N}(x_{i_n})] \ldots] =$$

$$= \max_{k \in \mathcal{M}^{i_n}} \Big\{ (\tilde{L}^s_{\mathrm{d}(i_n)} - \tilde{L}^s_{\mathrm{d}(k)}) p(k) f_k(x_{i_n}) +$$

$$+ \sum_{j_{n+2} \in \mathcal{M}^k} [(\tilde{L}^s_{\mathrm{d}(k)} - \tilde{L}^s_{\mathrm{d}(j_{n+2})}) q^*(j_{n+2} / k, j_{n+2}) \times$$

$$\times p(j_{n+2}) f_{j_{n+2}}(x_{i_n}) +$$

$$+ \ldots + \tilde{L}^s_{\mathrm{d}(j_{N-1})} \sum_{j_N \in \mathcal{M}^{j_{N-1}}} [q^*(j_N / k, j_N) \times$$

$$\times\, p(j_N)\, f_{j_N}(x_{i_n})]...]\Big\}$$

for $i_n \in \mathcal{M}(n)$, n=0,1,2,...,N–1, where $q^*(j_N / i_{n+1}, j_N)$ denotes probability of accurate classification of the object of the class j_N in further stages using π_N^* strategy rules on condition that on the n-th stage the i_{n+1} decision has been made.

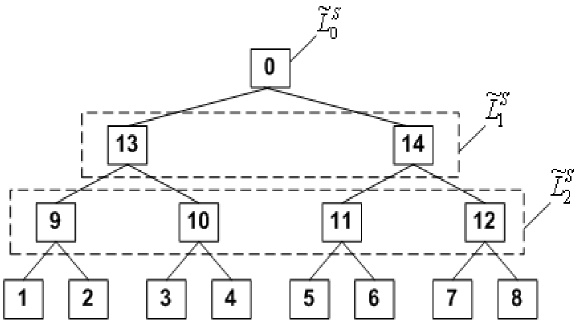

Fig. 1 Interpretation of stage-dependent fuzzy loss function

4 Project of simulation investigations

Let us consider the two-stage binary classifier present in Fig 2. Four classes have identical a priori probabilities which are equal 0.25. Class-conditional probability density functions of features $\boldsymbol{X_5}$ and $\boldsymbol{X_6}$ are following:

$$f_1(x_5) = \begin{cases} 4-|x_5-2|, & \text{while } 1.5 \le x_0 \le 2.5, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_2(x_5) = \begin{cases} 4-|x_5-2.5|, & \text{while } 2 \le x_0 \le 3, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_3(x_6) = \begin{cases} 4-|x_5-2|, & \text{while } 1.5 \le x_0 \le 2.5, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_4(x_6) = \begin{cases} 4-|x_6-1.3|, & \text{while } 0.8 \le x_0 \le 1.8, \\ 0 & \text{otherwise}, \end{cases}.$$

The feature $\boldsymbol{X}_0$ is triangular in each class with the following class-conditional probability density functions:

$$f_1(x_0) = f_2(x_0) = \begin{cases} 2-|x_5-1|, & \text{while } 0 \le x_0 \le 2, \\ 0 & \text{otherwise}, \end{cases}$$

$$f_3(x_0) = f_4(x_0) = \begin{cases} 2-|x_5-2|, & \text{while } 1 \le x_0 \le 3, \\ 0 & \text{otherwise}, \end{cases}$$

The stage-dependent fuzzy loss function are the following:

case 1 $\tilde{L}_0^S = (2, 3, 3, 4)_T$, $\tilde{L}_1^S = (1, 2, 2, 3)_T$,

case 2 $\tilde{L}_0^S = (2, 4, 4, 6)_T$, $\tilde{L}_1^S = (0, 0.5, 0.5, 1)_{Tr}$,

case 3 $\tilde{L}_0^S = (1, 1.5, 2.5, 3)_{Tr}$, $\tilde{L}_1^S = (0, 0, 0.5, 1)_{Tr}$,

case 4 $\tilde{L}_0^S = (1, 1.5, 2.5, 3)_{Tr}$, $\tilde{L}_1^S = (0.5, 1, 1,5)_T$,

and are described by triangular or trapezoidal membership function**.**

Due to the peculiar distribution of $\boldsymbol{X_5}$ and $\boldsymbol{X_6}$, the decision rules Ψ_5^* and Ψ_6^*, at the second stage of classification, the separation point are following: $x_5^{'} = 2.25$ and $x_6^{'} = 0.9$. Let us now determine the rule Ψ_0^* at the first stage of classification. From (7) we obtain:

$$\Psi_0^*(x_0) = \begin{cases} 5, \quad \text{if} & \begin{aligned} &(\tilde{L}_0^S - \tilde{L}_1^S)\, p(5)\, f_5(x_0) + \\ &+ \tilde{L}_1^S\,(q^*(1/5,1)\, p(1)\, f_1(x_0) + \\ &+ q^*(2/5,2)\, p(2)\, f_2(x_0)) > \\ &> (\tilde{L}_0^S - \tilde{L}_1^S)\, p(6)\, f_6(x_0) + \\ &+ \tilde{L}_1^S\,(q^*(3/6,3)\, p(3)\, f_3(x_0) + \\ &+ q^*(4/6,4)\, p(4)\, f_6(x_0)), \end{aligned} \\ 6 & \text{otherwise}, \end{cases}$$

Putting now the data of example, and use Campos-Gonzalez method for comparison fuzzy risk, we finally obtain results presented in Fig. 3-6, where value of point $x_0^{'}$ (separation point for decision regions at the first stage) in function of parameter λ is presented.

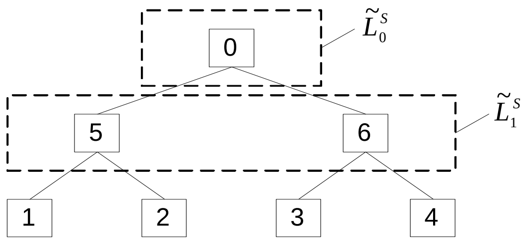

Fig. 2 Illustrate of simulation – decision tree

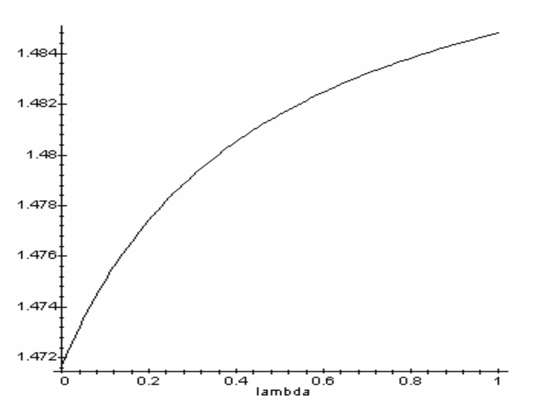

Fig. 3 Separation point for decision regions at the first stage – case 1

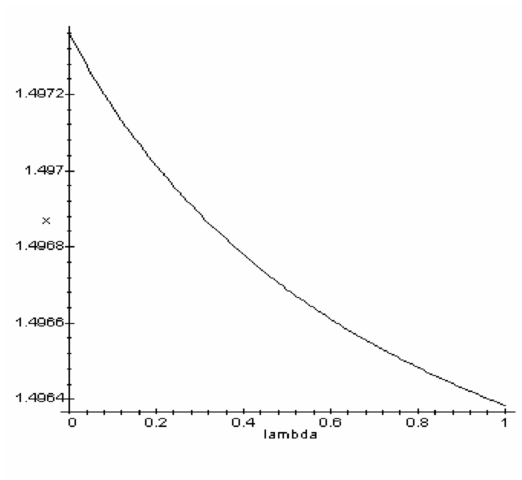

Fig. 4 Separation point for decision regions at the first stage – case 2

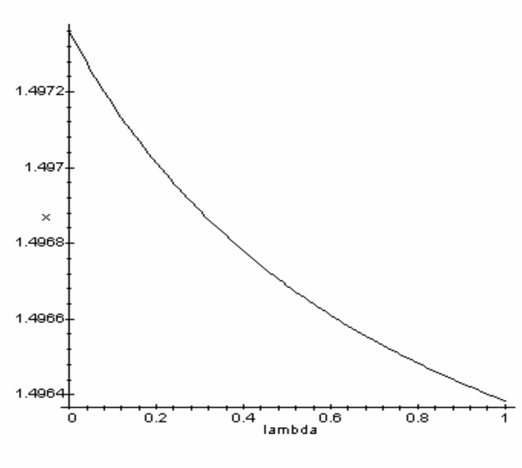

Fig. 5 Separation point for decision regions at the first stage – case 3

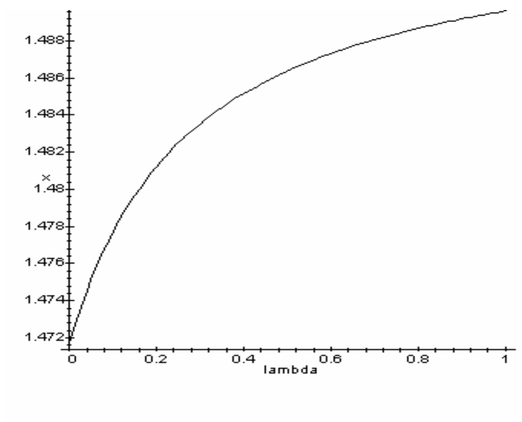

Fig. 6 Separation point for decision regions at the first stage – case 4

For crisp stage-dependent loss function (equal 2 and 1 on the first and the second stage respectively), we have x_0' =1.48.

5 Conclusions

In the paper we have presented the multistage Bayes classifier with a full probabilistic information. In this recognition model a priori probabilities of classes and class-conditional probability density functions are given. Additionally, consequences of wrong decision are fuzzy-valued and are represented by triangular or trapezoid fuzzy numbers. In this work we have considered algorithm for hierarchical classifier with stage-dependent fuzzy loss function when observation of the features are crisp. We use Campos-Gonzalez method for comparision fuzzy risk. Simulation investigations presented infuence of parameter λ on separation point for decision regions at the first stage. This paper contribute to a better understanding of the impact of the choice of a fuzzy numbers which describe stage-dependent loss function in multistage classifier.

Acknowledgements: This work is supported by The Polish State Committee for Scientifc Research under the grant which is realizing in years 2006-2009.

6 References

[1] S. Baas, and H. Kwakernaak. Rating and Ranking of Multi-Aspect Alternatives Using Fuzzy Sets. *Automatica*, 13:47-58, 1997.

[2] J. Berger. Statistical Decision Theory and Bayesian Analysis. Springer-Verlag. New York. 1993.

[3] G. Bortolan, and R. Degani. A Review of Some Methods for Ranking Fuzzy Subsets. *Fuzzy Sets and Systems* 15:1-19, 1985.

[4] L.M. Campos, and A. González. A Subjective Approach for Ranking Fuzzy Numbers. *Fuzzy Sets and Systems*, 29:145-153, 1989.

[5] M. Gil, and M. Lopez-Diaz. Fundamentals and Bayesian Analyses of Decision Problems with Fuzzy-Valued Utilities. *Int. J. of Approx. Reas.*, 15:203-224, 1996.

[6] M. Gil, and M. Lopez-Diaz. A Model for Bayesian Decision Problems Involving Fuzzy-Valued Consequences. Proc. 6-th Int. Conf. on Inf. Proc. and Manag. of Uncer. in Know. Based Systems, Granada pp.495—500

[7] R. Jain. Decision-Making in the Presence of Fuzzy Variables. *IEEE Trans. Sys. Man and Cyber*, 6:698-703, 1976.

[8] M. Kurzyński. On the Multistage Bayes Classifier. *Pattern Recognition*, 21:355-365, 1988.

[9] M. Kurzyński. Decision Rules for a Hierarchical Classifier. *Pattern Recognition Letters,* 1:305-310, 1983.

[10] M. Lopez-Diaz, and M.A. Gil. The λ-average value of the expected value of a fuzzy random variable. *Fuzzy Sets and Systems*, 99:347-352, 1996.

[11] R. Viertl. Statistical Methods for Non-Precise Data. CRC Press. Boca Raton. 1996.

[12] R. Yager. A procedure for Ordering Fuzzy Subsets of the Unit Interval. *Inf. Science*, 22:143-156, 1981.